



BIG DATA HADOOP AND SPARK TRAINING COURSE

altraSys Technologies

Learning partner for your team or your organization
Nurture talent with instructor-led courses

Sharing Big Thoughts

Today, business world is witnessing huge influx of data in exponential rate, virtually posing a challenge to all forward looking business leaders to leverage this corporate asset seating in form of Big Data. A recent study revealed that 85% of business leaders believe that big data will revolutionize business operations, in the same way the Internet did many years ago. The first step in leveraging the big data is to find a trained resource within to establish the ability to access and retrieve the useful information to support decision making and day-to-day operations. Obviously a gap is visible that the big data analyst is required with proper skill to bridge the gap between information, insights and action. Are you ready to fill the gap?

The altraSys Big Data team has brought to you a course design to train in the basic to advance level according to your learning plan. The team encourages you to go through course details below and enroll for the altraSys flexible training schedule.



altraSys is committed to address the growing requirements of the industry by providing the best-in-class training that empowers professionals to approach real-world scenarios

Course Details

BIG DATA HADOOP AND SPARK TRAINING COURSE

About The Course

We offer a full-fledged hands-on Big Data Hadoop and Spark training designed by the industry experts to build your in-depth knowledge of **Big Data Hadoop ecosystem** and computing framework using **Hadoop Map Reduce** and **Spark** including **HDFS, YARN, Sqoop, Flume, Pig, Hive, Impala, HBase, Kafka, Oozie and ZooKeeper**. The training emphasizes on use cases to teach each component, case study practicals and end-to-end workflows. At the end, the participants are equipped to be readiness to appear successfully the Cloudera/Hortonworks OEM certifications.



Why you should learn “Big Data”?

The world is heading towards digitalization which means Big Data is here to stay. In fact, the importance of Big Data Analytics is going to continue growing in these coming years. Business leaders across all sectors are grappling with the strategic implications of Big Data Analytics for their organizations and industry ecosystems. Now it is the perfect time to learn and adopt the Big Data Technology and bring a qualitative transformation within you and your organization.

“Big data is at the foundation of all of the megatrends that are happening today, from social to mobile to the cloud to gaming.”

Course Details

BIG DATA HADOOP AND SPARK TRAINING COURSE

What Makes Data, “BIG” Data?

With every single passing day, Data volume is increasing exponentially contributed by countless email exchanges, smart meters, mobile devices, social media chats, RFID, web logs besides usual business operations. Data in unstoppable velocity, in different varieties come pouring in with bigger challenges, opportunities and actionable insights hidden beneath the voluminous heap that gather no moss.

Message to Candidates - Is Big Data Hadoop for ‘U’?

All forward looking organizations need a big data strategy and solution developer to understand tomorrow’s market trends to position themselves on competitive advantages. Today there has been a critical gap between demand and availability of trained hands to analyze the big data by entering enterprise Big Data Hadoop eco-system on daily basis. It has become a truth and fact of the life for corporate leaders to struggle to hire talents to manage and overcome the shortage of trained entering enterprise Hadoop eco-system on daily basis. Opting Hadoop Training can turn out to be a promising opportunity for you to spice up your career and organizations. Hadoop certification is targeted for professional aspiring to accelerate career growth in Big Data analytics. **Why not it is ‘U’?**

FEATURES OF THIS TRAINING COURSE



Instructor Led Training



Application Based Learning



Interactive Session



Industry Specific Projects



Dedicated Support Team



Cloud Lab Access

Course Details

BIG DATA HADOOP AND SPARK TRAINING COURSE

Commercial Pricing

Decision on price model is currently deferred till such time training duration and course coverage are discussed and decided.

Certification & Training Support

This training course content has been so deigned to align with OEM certification curriculum.

Mode of Conduct of Training

- ✚ Instructor-Led Online Training
- ✚ Classroom Training
- ✚ Corporate/Special Training

altraSys Cloud Lab Access

During training period you will be provided a big data cloud lab for code practicing. Additional 30 days altrasyS cloud lab access will be provided after the completion of training.

Details of our Regular & Advanced Course

**BIG DATA HADOOP & SAPRK
REGULAR TRAINING
(72 hours)**

**BIG DATA HADOOP & SAPRK
ADVANCED TRAINING
(45 hours)**

Course Content

BIG DATA HADOOP AND SPARK TRAINING COURSE

BIG DATA HADOOP & SAPRK **REGULAR** TRAINING

Our Big Data Hadoop & Spark Regular training course is designed to cover all required big data tools. This course will help you to understand the basics and advanced concept of Hadoop & Spark with all components like HDFS, Map Reduce, YARN, Sqoop, Flume, Hive, Impala, Spark Core API, Spark SQL, Spark Streaming, Oozie, ZooKeeper and some basics of Hadoop administration. Most importantly, this course focused on hands-on exercises, real-time use cases and topic wise code practice which will help you to get practical use of tools and codes rather than knowing only theoretical concepts.

Course Duration: 72 Hours

Modules:

01. **Big Data and Hadoop Ecosystem** (02 Hours)
02. **Hadoop Cluster and HDFS** (03 Hours)
03. **Hadoop Map Reduce and YARN** (03 Hours)
04. **Apache Sqoop** (08 Hours)
05. **Apache Flume** (06 Hours)
06. **Apache Pig** (04 Hours)
07. **Apache Hive** (09 Hours)
08. **Apache Impala** (03 Hours)
09. **Apache Spark Using Scala** (18 Hours)
10. **Oozie & Zookeeper** (03 Hours)
11. **Hadoop Installation** (09 Hours)
12. **Real-time Projects & Assignments** (10 Hours)

Training Pre-requisites:

There are no such prerequisites for Big Data Hadoop & Spark Regular Training Course. However, the learning path becomes smooth if the participants have previous exposure to UNIX, SQL and Java. It is not mandatory but if you wish to brush up the required skills we can offer you a 10 hours basic course on “Linux, SQL and Java”.

Course Content

BIG DATA HADOOP AND SPARK TRAINING COURSE

Course Curriculum of “Big Data Hadoop & Spark Regular Training Course”:

Module 1: Introduction to Big Data and Hadoop Ecosystem

Topic.1.1: Introduction to Big Data

- Overview to Big Data and Hadoop
- Traditional systems and associated problems
- Necessity of Big Data and Hadoop in the industry
- Various implementations of Big Data
- Future of Big Data

Topic.1.2: Hadoop Ecosystem

- Big Data Hadoop Framework
- Components of the Hadoop ecosystem
- Hadoop Flavours – Apache, Cloudera, Hortonworks, and more

Module 2: Hadoop Framework and HDFS

Topic.2.1: Hadoop Framework

- Introduction to Hadoop
- Why Hadoop is important?
- Features of Hadoop
- Hadoop Architecture
- Hadoop Pros and Cons

Topic.2.2: Hadoop Distributed File System (HDFS)

- What is HDFS (Hadoop Distributed File System)?
- HDFS Master-Slave Architecture
- Master Node and Slave Node
- NameNode and DataNode
- Data Block in HDFS
- Replication Management
- Rack Awareness
- High Availability
- Federation
- Disk Balancer

Course Content

BIG DATA HADOOP AND SPARK TRAINING COURSE

Topic2.3: Hadoop Cluster

- Federation and High Availability Architecture
- Hadoop Cluster Modes
- Distributed Cache
- Basic Hadoop Shell Commands
- Commissioning and Decommissioning Nodes

Topic.2.4: Understanding HDFS Commands & Web UI

- Hadoop HDFS Web UI and HDFS explorer
- Various HDFS Commands and Operations

Module 3: Hadoop Map Reduce and YARN Framework

Topic.3.1: Map Reduce – The Processing Layer

- What is Map Reduce?
- Traditional vs Map Reduce Approach
- Map Reduce Architecture
- Detailed understanding of the working of Map Reduce
- The Mapping and Reducing Phase
- Working of Driver, Combiners, Partitioners, Input/Output Formats, Shuffle and Sort.
- Anatomy of Map Reduce Program
- Input Splits, Relation between Input Splits and HDFS Blocks
- Map Reduce Job Configuration & Submission
- Task Execution & Environment

Topic.3.2: Hadoop YARN Framework – Resource Management

- What is YARN?
- YARN Architecture
- YARN Components
- ResourceManager
- NodeManager
- YARN Map Reduce Application Execution Flow

Course Content

BIG DATA HADOOP AND SPARK TRAINING COURSE

Module 4: Apache SQOOP

Topic.4.1: Overview of Sqoop

- Introduction to Sqoop
- Need of Sqoop and Sqoop Features

Topic.4.2: Working with Sqoop Tools

- Sqoop Architecture
- Sqoop Eval and Codegen tools
- Sqoop Import and Export tools
- Sqoop List Databases and Table tools
- File Formats
- Sqoop with Hive

Topic.4.3: Sqoop Jobs

- Sqoop Jobs
- Saved Jobs and Incremental Imports

Topic.4.4: Sqoop Configurations

- Sqoop Performance Optimization By Controlling Parallelism
- Specific Connectors

Module 5: Apache Flume

Topic.5.1: Overview of Flume

- Introduction to Flume
- Why Flume, Flume Features

Topic.5.2: Working with Flume

- Flume Architecture
- Source, Sink and Channel
- Data Flow Model
- Setting up an Agent, Fetching Data
- Data Transfer to HDFS
- Configuring multi-agent flow
- Different Flume Sources(Avro, Thrift, Exec, JMS)
- Kafka Source

Course Content

BIG DATA HADOOP AND SPARK TRAINING COURSE

Module 5: Apache Flume

Topic.5.1: Overview of Flume

- Introduction to Flume
- Why Flume
- Flume Features

Topic.5.2: Working with Flume

- Flume Architecture
- Source, Sink and Channel
- Data Flow Model
- Setting up an Agent
- Fetching Data
- Data Transfer to HDFS
- Configuring multi-agent flow
- Different Flume Sources(Avro, Thrift, Exec, JMS)
- Kafka Source

Module 6: Apache Pig

Topic.6.1: Overview of Pig

- Introduction to Apache Pig
- Need for Pig
- Features of Pig
- Pros and Cons of Pig

Topic.6.2: Working with Pig

- Pig Architecture
- Pig Components & Pig Execution
- Pig Data Types & Data Models in Pig
- Pig Latin Programs
- Shell and Utility Commands
- Pig UDF & Pig Streaming

Course Content

BIG DATA HADOOP AND SPARK TRAINING COURSE

Module 7: Apache Hive

Topic.7.1: Overview of Hive

- Introduction to Apache Hive
- Importance of Hive
- Comparison with Traditional Database
- Features and Limitations of Hive

Topic.7.2: Understanding Hive

- Hive Architecture and Components
- Hive Metastore
- Hive Data Models/Units
- Hive – View and Index
- Hive Partition and Bucketing
- Hive Data Types
- Hive Transactions
- Hive Configuration Properties

Topic.7.3: Hive Language

- Hive Commands and CLIs
- File Formats and Compression
- Hive Data Definition Language
- Hive Tables (Managed Tables and External Tables)
- Hive Data Manipulation Language
- Data Retrieval: Queries
- Joins

Topic.7.4: Hive Advanced

- Hive SerDe
- Hive Clients
- Hive Web Interface (Query through HUE)
- Hive Built-In Functions and UDFs

Topic.7.5: Hive Comparison

- Hive Vs HBase
- Hive Vs Pig
- Hive Vs Impala

Course Content

BIG DATA HADOOP AND SPARK TRAINING COURSE

Module 8: Apache Impala

Topic.8.1: Overview of Impala

- Introduction to Impala
- Why Use Impala

Topic.8.2: Working with Impala

- Impala Architecture
- Impala Server Components
- Impala Shell & Basic Commands
- How Impala uses HDFS, Hive and HBase
- Impala Shell, ODBC & JDBC and Hue
- Database, Table and Clause Operations
- Partitioning & Logging
- Build-In Functions and UDFs
- Dealing with Parquet Files with Unknown Schema

Module 9: Apache Spark Using Scala

Topic.9.1: Overview of Spark

- What is Spark
- Spark Ecosystem
- Spark Components
- Importance of Spark
- Features and Limitation of Spark

Topic.9.2: Understanding Spark Environment

- Spark Architecture
- How to Use Scala or Python
- Spark Context
- Spark Stage
- Spark Executor
- Spark Configuration and Parameters
- Use of sbt

Course Content

BIG DATA HADOOP AND SPARK TRAINING COURSE

Topic.9.3: Spark Core API

- RDD, RDD Features
- Ways to Create RDD
- Reading Different File Formats
- Paired RDD
- RDD Transformations and Actions
- Row Level Transformations
- RDD Aggregations
- Joining RDD
- Saving RDD to a file
- RDD Lineage
- RDD Limitations

Topic.9.4: Spark SQL

- Spark SQL Introduction
- SQL Context
- Spark DataFrames
- Spark Datasets
- RDD vs DataFrames vs Datasets
- DataFrame Operations
- Registering Temp Table
- Aggregate Functions
- Analytics Functions
- Windowing Functions
- Schema manual inferring, working with CSV files, reading of JDBC tables
- Writing Saprk SQL Applications
- Creating HiveContext
- Writing Data Frame to Hive
- Deploying Hive on Spark as the execution engine

Topic.9.5: Spark Streaming

- Getting Started with Spark Streaming
- Data Structure (DStream) and APIs
- DStream Data Transformations
- Check pointing in Spark Streaming
- Integrating with Flume
- Integrating with Kafka

Course Content

BIG DATA HADOOP AND SPARK TRAINING COURSE

Module 10: Oozie & Zookeeper

Topic.10.1: Overview of Oozie

- Introduction to Oozie
- Oozie Components
- Oozie Workflow
- Why Oozie

Topic.10.1: Overview of Zookeeper

- Zookeeper Overview
- Zookeeper Architecture
- Zookeeper Workflow
- Why Zookeeper

Module 11: Hadoop Administration Essentials

Topic.11.1: Setup and Installation of Single-Node and Multi-Node Hadoop Cluster

- Hadoop environment setup and pre-requisites
- Hadoop Installation
- Hadoop Configuration
- Various Hadoop Installation Modes

Module 12: Projects & Assignments

- You will be working on different real-life use cases to learn the industrial use of Hadoop components like Map Reduce, Sqoop, Flume, Pig, Hive, Spark and Spark Streaming.

Course Content

BIG DATA HADOOP AND SPARK TRAINING COURSE

BIG DATA HADOOP & SPARK **ADVANCED** TRAINING

Our Big Data Hadoop & Spark advanced training course is designed as a project driven training course. In this course we will brush up your knowledge on the big data tools and you will be learning the real use of big data tools by working on different types of real-life use cases across all industries. More advanced tools like HBase, Kafka, and Spark MLib, Spark GraphX are added in this course curriculum. You will be also learning the Hadoop administration essentials in a deeper way.

The main objective of this course is to make you industry ready by providing you a series of real-life projects on different industry, on different tools. It will help you to understand the concept as well as the practical use of all components and the solution architecture. It is designed to give you the complete knowledge of big data platform so that you can start working on big data project as soon as completing this training course.

Course Duration: 45 Hours

01. **Big Data and Hadoop Ecosystem** (02 Hours)
02. **Hadoop Cluster and HDFS** (01 Hour)
03. **Hadoop Map Reduce and YARN** (01 Hour)
04. **Apache Sqoop** (01 Hours)
05. **Apache Flume** (01 Hours)
06. **Apache Pig** (01 Hours)
07. **Apache Hive** (02 Hours)
08. **Apache Impala** (01 Hours)
09. **HBase** (06 Hours)
10. **Apache Spark (with Machine Learning & GraphX)** (10 Hours)
11. **Apache Kafka** (04 Hours)
12. **Oozie & Zookeeper** (01 Hours)
13. **Hadoop Administration Essentials** (04 Hours)
14. **Real-time Projects & Assignments** (10 Hours)

Pre-requisites:

- The participants must have completed our “Big Data Hadoop & Spark Training Course” or else they must have the knowledge of all the big data tools to start our “Advanced Training Course”.
- The learning path becomes smooth if the participants have previous exposure to UNIX, SQL and Java. However, we offer a 10 hours basic course on “Linux, SQL and Java” to brush up the required skills.

Course Content

BIG DATA HADOOP AND SPARK TRAINING COURSE

Course Curriculum of “Big Data Hadoop & Spark Advanced Training Course”:

Module 1: Introduction to Big Data and Hadoop Ecosystem

Topic.1.1: Introduction to Big Data

- Overview to Big Data and Hadoop
- Traditional systems and associated problems
- Necessity of Big Data and Hadoop in the industry
- Various implementations of Big Data
- Future of Big Data

Topic.1.2: Hadoop Ecosystem

- Big Data Hadoop Framework
- Components of the Hadoop ecosystem
- Hadoop Flavours – Apache, Cloudera, Hortonworks, and more

Module 2: Hadoop Framework and HDFS

Topic.2.1: Hadoop Framework

- Introduction to Hadoop
- Why Hadoop is important?
- Features of Hadoop
- Hadoop Architecture
- Hadoop Pros and Cons

Topic.2.2: Hadoop Distributed File System (HDFS)

- What is HDFS (Hadoop Distributed File System)?
- HDFS Master-Slave Architecture
- Master Node and Slave Node
- NameNode and DataNode
- Data Block in HDFS
- Replication Management
- Rack Awareness
- High Availability
- Federation
- Disk Balancer

Course Content

BIG DATA HADOOP AND SPARK TRAINING COURSE

Topic2.3: Hadoop Cluster

- Federation and High Availability Architecture
- Hadoop Cluster Modes
- Distributed Cache
- Basic Hadoop Shell Commands
- Commissioning and Decommissioning Nodes

Topic.2.4: Understanding HDFS Commands & Web UI

- Hadoop HDFS Web UI and HDFS explorer
- Various HDFS Commands and Operations

Module 3: Hadoop Map Reduce and YARN Framework

Topic.3.1: Map Reduce – The Processing Layer

- What is Map Reduce?
- Traditional vs Map Reduce Approach
- Map Reduce Architecture
- Detailed understanding of the working of Map Reduce
- The Mapping and Reducing Phase
- Working of Driver, Combiners, Partitioners, Input/Output Formats, Shuffle and Sort.
- Anatomy of Map Reduce Program
- Input Splits, Relation between Input Splits and HDFS Blocks
- Map Reduce Job Configuration & Submission
- Task Execution & Environment

Topic.3.2: Hadoop YARN Framework – Resource Management

- What is YARN?
- YARN Architecture
- YARN Components
- ResourceManager
- NodeManager
- YARN Map Reduce Application Execution Flow

Course Content

BIG DATA HADOOP AND SPARK TRAINING COURSE

Module 4: Apache SQOOP

Topic.4.1: Overview of Sqoop

- Introduction to Sqoop
- Need of Sqoop
- Sqoop Features

Topic.4.2: Working with Sqoop Tools

- Sqoop Architecture
- Sqoop Eval and Codegen tools
- Sqoop Import and Export tools
- Sqoop List Databases and Table tools
- File Formats
- Sqoop with Hive

Topic.4.3: Sqoop Jobs

- Sqoop Jobs
- Saved Jobs and Incremental Imports

Topic.4.4: Sqoop Configurations

- Sqoop Performance Optimization By Controlling Parallelism
- Specific Connectors

Module 5: Apache Flume

Topic.5.1: Overview of Flume

- Introduction to Flume
- Why Flume
- Flume Features

Topic.5.2: Working with Flume

- Flume Architecture
- Source, Sink and Channel
- Data Flow Model
- Setting up an Agent
- Fetching Data

Course Content

BIG DATA HADOOP AND SPARK TRAINING COURSE

- Data Transfer to HDFS
- Configuring multi-agent flow
- Different Flume Sources(Avro, Thrift, Exec, JMS)
- Kafka Source

Module 6: Apache Pig

Topic.6.1: Overview of Pig

- Introduction to Apache Pig
- Need for Pig
- Features of Pig
- Pros and Cons of Pig

Topic.6.2: Working with Pig

- Pig Architecture
- Pig Components & Pig Execution
- Pig Data Types & Data Models in Pig
- Pig Latin Programs
- Shell and Utility Commands
- Pig UDF & Pig Streaming

Module 7: Apache Hive

Topic.7.1: Overview of Hive

- Introduction to Apache Hive
- Importance of Hive
- Comparison with Traditional Database
- Features and Limitations of Hive

Topic.7.2: Understanding Hive

- Hive Architecture and Components
- Hive Metastore
- Hive Data Models/Units
- Hive – View and Index
- Hive Partition and Bucketing
- Hive Data Types

Course Content

BIG DATA HADOOP AND SPARK TRAINING COURSE

- Hive Transactions
- Hive Configuration Properties

Topic.7.3: Hive Language

- Hive Commands and CLIs
- File Formats and Compression
- Hive Data Definition Language
- Hive Tables (Managed Tables and External Tables)
- Hive Data Manipulation Language
- Data Retrieval: Queries
- Joins

Topic.7.4: Hive Advanced

- Hive SerDe
- Hive Clients
- Hive Web Interface (Query through HUE)
- Hive Built-In Functions and UDFs

Topic.7.5: Hive Comparison

- Hive Vs HBase
- Hive Vs Pig
- Hive Vs Impala

Module 8: Apache Impala

Topic.8.1: Overview of Impala

- Introduction to Impala
- Why Use Impala

Topic.8.2: Working with Impala

- Impala Architecture
- Impala Server Components
- Impala Shell & Basic Commands
- How Impala uses HDFS, Hive and HBase
- Impala Shell, ODBC & JDBC and Hue
- Database, Table and Clause Operations
- Partitioning & Logging
- Build-In Functions and UDFs
- Dealing with Parquet Files with Unknown Schema

Course Content

BIG DATA HADOOP AND SPARK TRAINING COURSE

Module 9: HBase

Topic.9.1: Overview of HBase

- What is HBase
- NoSQL Databases and its types
- HBase vs HDFS
- HBase v/s RDBMS
- Features and Limitation of HBase

Topic.9.2: Working with HBase

- HBase Components
- HBase Architecture
- HBase Run Modes
- HBase Configuration
- HBase Cluster Deployment
- HBase Data Model
- HBase Shell
- HBase Client API
- Hive Data Loading Techniques
- HBase Bulk Loading
- Getting and Inserting Data
- HBase Filters

Module 10: Apache Spark Using Scala

Topic.10.1: Overview of Spark

- What is Spark
- Spark Ecosystem
- Spark Components
- Importance of Spark
- Features and Limitation of Spark

Topic.10.2: Understanding Spark Environment

- Spark Architecture
- How to Use Scala or Python
- Spark Context
- Spark Stage

Course Content

BIG DATA HADOOP AND SPARK TRAINING COURSE

- Spark Executor
- Spark Configuration and Parameters
- Use of sbt

Topic.10.3: Spark Core API

- RDD
- RDD Features
- Ways to Create RDD
- Reading Different File Formats
- Paired RDD
- RDD Transformations and Actions
- Row Level Transformations
- RDD Aggregations
- Joining RDD
- Saving RDD to a file
- RDD Lineage
- RDD Limitations

Topic.10.4: Spark SQL

- Spark SQL Introduction
- SQL Context
- Spark DataFrame
- Spark Datasets
- RDD vs DataFrame vs Dataset
- DataFrame Operations
- Registering Temp Table
- Aggregate Functions
- Analytics Functions
- Windowing Functions
- Schema manual inferring, working with CSV files, reading of JDBC tables
- Writing Saprk SQL Applications
- Creating HiveContext
- Writing Data Frame to Hive
- Deploying Hive on Spark as the execution engine

Topic.10.5: Spark Streaming

- Getting Started with Spark Streaming
- Data Structure (DStream) and APIs
- DStream Data Transformations
- Check pointing in Spark Streaming

Course Content

BIG DATA HADOOP AND SPARK TRAINING COURSE

- Integrating with Flume
- Integrating with Kafka

Topic.10.6: Machine Learning with Spark MLlib

- Why Machine Learning?
- What is Machine Learning?
- Where Machine Learning is used?
- Introduction to MLlib
- Features of MLlib and MLlib Tools
- Various ML algorithms supported by MLlib
- Supervised Learning - Linear Regression, Logistic Regression, Decision Tree, Random Forest
- Unsupervised Learning - K-Means Clustering & How It Works with MLlib
- Analysis on US Election Data using MLlib (K-Means)

Topic.10.7: Spark GraphX

- Spark GraphX Introduction
- Spark GraphX Features
- Key concepts of Spark GraphX programming
- Spark GraphX - Operators
- Spark GraphX - Algorithms
- Use Cases of Spark GraphX

Module 11: Apache Kafka

Topic.11.1: Overview of Kafka

- Introduction to Kafka
- Need for Kafka
- Kafka Features
- Kafka Concepts

Topic.11.2: Working with Kafka

- Kafka Architecture
- Kafka Components
- Work of ZooKeeper
- Kafka Cluster
- Types of Kafka Clusters
- Kafka Producer
- Kafka Consumer

Course Content

BIG DATA HADOOP AND SPARK TRAINING COURSE

- Kafka Internals
- Kafka Monitoring and Kafka Connect
- Kafka Stream Processing
- Integration of Kafka With Hadoop, Flume and Spark

Module 12: Oozie & Zookeeper

Topic.12.1: Overview of Oozie

- Introduction to Oozie
- Oozie Components
- Oozie Workflow
- Why Oozie

Topic.12.2: Overview of Zookeeper

- Zookeeper Overview
- Zookeeper Architecture
- Zookeeper Workflow
- Why Zookeeper

Module 13: Hadoop Administration Essentials

Topic.13.1: Setup and Installation of Single-Node and Multi-Node Hadoop Cluster

- Hadoop environment setup and pre-requisites
- Hadoop Installation
- Hadoop Configuration
- Various Hadoop Installation Modes

Module 14: Real-time Projects & Assignments

- You will be working on different real-life use cases to learn the industrial use of Hadoop components like Map Reduce, Sqoop, Flume, Pig, Hive, Spark and Spark Streaming with Kafka.